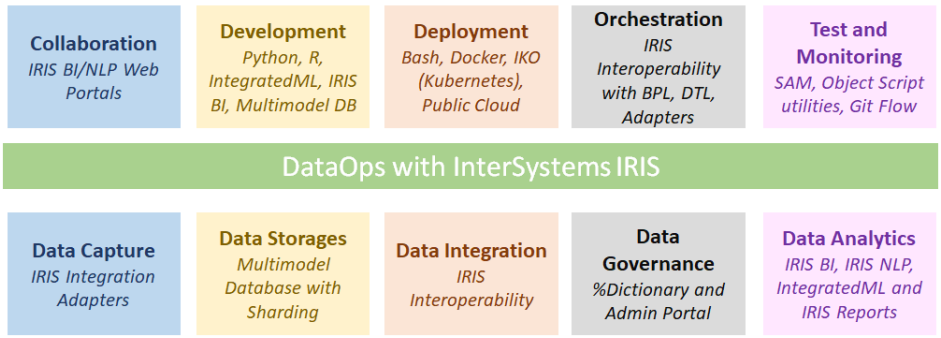
文章
Louis Lu · 四月 15, 2021 阅读大约需 6 分钟

基于InterSystems IRIS 的数据运维DataOps

InterSystems IRIS 下使用 DataOps



Gartner 对 DataOps 的定义是：“ DataOps
是一种协作式的数据管理方法，侧重于改善整个组织中数据管理者和数据消费者之间数据流的沟通、整合与自动化。 DataOps 的目标是创建可预测的数据、数据模型和相关项目的交付和变更管理，从而更快地交付价值。 DataOps 采取特殊技术手段和相应治理水平自动化数据交付的设计、部署和管理，以元数据提高动态环境中数据的易用性和价值。 ”

2014 年 6 月 19 日，InformationWeek 特约编辑 Lenny Liebmann 发表于 IBM Big Data & Analytics Hub 的题为“ 3 reasons why DataOps is essential for big data success ”的文章中首次提出 DataOps 这一概念。 DataOps 一词后被 Andy Palmer 推广到 Tamr。 DataOps 是“ 数据运营 ”的专属名称。 2017 年对 DataOps 来说是意义重大的一年，生态系统取得巨大发展，分析师覆盖范围进一步扩张，关键字搜索量以及调查、出版物和开源项目数均有所提升。 Gartner 在 2018 年的 Hype Cycle for Data Management 中添加了 DataOps 。
（资料来源：https://en.wikipedia.org/wiki/DataOps）

DataOps 宣言确立了以下 DataOps 原则：(https://www.dataopsmanifesto.org/dataops-manifesto.html)

1. **持续满足客户需求**
   ：我们的首要任务是在几分钟到几周内及尽早并持续交付有价值的内容给客户，并以此满足客户的需求。
2. **有价值工作的分析**
   ：我们认为评价工作效率的主要度量指标是，提交了多少深度分析的内容、产出多少高准确度的数据以及在顶层框架和系统中贡献了多少。
3. **拥抱变化**：我们并不抗拒客户需求的变化，事实上，我们欣然接受这些变化，并以此产生竞争优势。
   我们相信，与客户直接交谈是最高效、最实用、最敏捷的沟通方式。
4. 这是一项团队运动：分析团队将始终具有各种角色、技能、偏好工具和头衔。
   多元化的背景和意见可以提高创新力和生产力。
5. **日常互动**：客户、分析团队和运营必须每天都在整个项目中协同工作。
6. **自发组织**：我们相信，最好的分析见解、算法、架构、需求和设计都来自于自发组织的团队。
7. **减少英雄主义**
   ：对于分析的深度和广度，需求正在加速扩大，我们认为分析团队应尽力减少英雄主义，创建可持续且可扩展的数据分析团队和流程。
8. **反思**：分析团队应定期根据客户、自身和运营统计提供的反馈意见开展自我反思，优化运营绩效。
9. **分析即代码**：分析团队使用各种工具对数据进行访问、集成、建模和可视化。
   从根本上讲，每种工具都会生成代码和配置，这些都会对数据进行操作，从而为进一步理解数据提供帮助。
10. **编排**：从头到尾编排数据、工具、代码、环境以配合分析团队的工作，这是分析成功的关键因素。
11. 使其可复现：结果需要可复现，我们对所有内容进行版本控制：包括数据、底层硬件、软件配置以及特定于工具链中各工具的代码和配置。
12. **一次性的工作环境**
    ：我们认为，必须为分析团队成员提供易于创建、隔离、安全并能反映其生产环境的一次性工作环境，从而

最大程度地降低分析团队的实验成本。

13. **简洁性**
：我们认为，持续关注卓越的技术和良好的设计不仅可以提高敏捷性，也可以提高简洁性，特别是突出显示未完成的工作量。
14. **分析即生产**：分析管道类似于精益生产线。 我们认为，DataOps 的一个基本概念是注重过程思维，提升分析及生产的效率。
15. **质量至上**：分析管道在基础上应该有能力自动检测代码、配置和数据中的异常 (jidoka) 与安全问题，并应向操作员提供持续反馈以避免错误 (poka yoke)。
16. **监视质量和性能**
：我们的目标是对性能、安全和质量措施进行持续监视，及时发现意外变化并生成运营统计信息。
17. **重用**：我们认为，提升分析及生产效率的一个基本方面是避免个人或团队重复以前的工作。
18. **缩短周期时间**
：从将客户需求转化为分析思路，到在开发中创建、并作为可重复生产的流程发布，再到最后的重构和重用产品，我们应尽全力将这一周期耗费的时间和精力降到最低。

当您分析这些原则时，也许会发现 InterSystems IRIS 能够在某些方面起到作用：

- 持续满足客户需求：您可以通过冲刺或迭代创建新的短集成产品、编排、IRIS 多维数据集、报告、BI 可视化和 ML 模型。
- 有价值工作的分析：IRIS 帮助您提供高质量的数据（在持久化类中使用production、适配器和类方法），并使您能够在 IRIS BI 数据透视表（分析设计器）和 IRIS NLP（文本分析）中进行数据探索。
- 自组织：IRIS 简化了自组织，借助统一的数据平台，您只需一个工具即可收集、处理、分析和发布见解。
- 反思：您可以通过此用户门户与用户互动并收集反馈，以改进交付的产品。
- 分析即代码：在 IRIS 数据模型中，多维数据集、仪表板都是代码，具有版本控制和治理功能。
- 编排：IRIS 数据平台可在单个工具中编排数据的引入、扩充、分析工作、数据可视化和 ML。
- 使其可复现：IRIS 支持使用 Docker、Kubernetes (IKO) 和 DevOps 复现结果。
- 一次性环境：IRIS 支持为集成、数据模型、BI 多维数据集和可视化创建 Docker 一次性环境。
- 简洁性：IRIS 数据多维数据集的创建非常简单，无需创建 ETL 脚本，分析器、多维数据集、仪表板的创建均实现可视化和网络化，并且可由用户而不仅是开发者团队完成。 同时，IntegratedML 允许在没有源代码开发的情况下针对常见场景创建 ML。
- 监视质量和性能：IRIS 使用 SAM 监视性能并具有 Web 管理门户。
- 重用：在 IRIS 中，DataOps 项目是类，这些类默认可扩展、可重用。
- 缩短周期时间：用户可以通过自助服务创建仪表板、分析、报告，以及发布和共享工作。

ODSC (https://opendatascience.com/maximize-upstream-dataops-efficiency-through-ai-and-ml-to-accelerate-analytics/) 指出以下 DataOps 策略：

InterSystems IRIS 对以上几点均有所帮助：

- 自助服务配置：用户可以创建和发布多维数据集与仪表板。
- 共享、标记、注解：用户门户可用于共享仪表板，IRIS Analytical Web Portal 允许用户创建、记录、整理到文件夹并标记您的工作。
- 扩充：BPL 可用于扩充数据。
- 准备：BPL、DTL、适配器和 ObjectScript 逻辑可以帮助准备数据。
- 数据市场：数据资产可以发布到 REST API 并通过 IRIS API Manager 获利。
- 数据目录：IRIS 中的数据被组织成类，这些类被存储在类目录系统 (%Dictonary) 中
- 配置文件与分类：可在用户门户和管理门户中为分析项目创建组、文件夹。
- 质量：IRIS 具有实用工具类，可生成示例数据和进行单元测试。
- 沿袭：在 IRIS 中，所有数据资产都相互连接，您可以从数据模型构建多维数据集，再从多维数据集构建仪表板，所有数据资产均可由数据管理者（IRIS 权限系统）控制。
- 掌控：通过管理门户，您可以掌控分析项目的各个方面。
- 数据库数据、文件数据、SaaS API、流：IRIS 为多模型，支持持久性以及数据和文本分析 (NLP)。 以 IRIS API Manager 支持 SaaS API，以 Integration Adapters 和 PEX（带有 Kafka）与 Streams 结合使用。
- 应用程序、BI 工具、分析沙盒：通过 IRIS，您可以使用您喜欢的语言（Java、Python、.NET、Node.js、ObjectScript）创建 DataOps 应用。

> 虽然 IRIS 是 BI 工具，但是在这个工具中，您可以将连接器与 Power BI 或 MDX 桥结合使用，将 IRIS 作为分析沙盒。

参见我反映 IRIS 和 DataOps 的汇总：

#分析 #InterSystems IRIS

源 URL:https://cn.community.intersystems.com/post/%E5%9F%BA%E4%BA%8Eintersystems-iris-%E7%9A%84%E6%95%B0%E6%8D%AE%E8%BF%90%E7%BB%B4dataops